23 January 1970

25X1A

MEMORANDUM FOR: [redacted] Chief DDP/SG

SUBJECT : Evaluation of S/360 FFS as an alternative for GICS Conversion

REFERENCE : Your memorandum of 9 January 1970 on the same subject.

25X1A

1. [redacted] has ascertained that NMCS maintains large tape files, in addition to other random access files, under S/360 FFS. He indicated Mr. Glen Stevener, Deputy for the Operations Group (OX-5-2935) is the best contact. [redacted] will contact you in the near future to discuss the possibility of visiting the NMCS installation. 25X1A

2. I concur with the statement in your memorandum that the most significant disadvantage in using S/360 FFS would be the apparent space requirements. The storage space explosion is caused by two related problems.

A. The necessity to format each field to a fixed length sufficient to hold the maximum length specified. Thus, using normal storage techniques, the [redacted] field which has a maximum length of 22 characters and occurs approximately 35,000 times in a particular file would require a storage allocation of 770,000 eight bit bytes. 25X1A

B. The necessity to leave "holes" in a formatted record when fields do not exist for that record. For example, if a periodic set consisted of three fields (say city, state and country) the numbers of full periodic subsets required is equivalent to the maximum occurrences of any of the fields. Country might be present 300 times and for simplicity let city and state be present 200 times. Then there would be 100 "holes" for each of the other two fields allocated but unused.

3. While examining the distribution by size for the [redacted] field it occurred to me that a significant reduction in storage requirements could be achieved utilizing an encoding scheme. The character distribution for the [redacted] field showed the average field length to be eight characters whereas the maximum is 22. If we are only concerned 25X1A

25X1A

with alpha characters (26). Then at most we need 5 bits to represent an alpha character. This would permit 33 percent reduction in field size and a corresponding reduction of 33 percent in the total storage required in an FFS implementation. This paragraph includes some over simplifications of the total problem but the facts of compaction indicate more analysis is desirable.

4. There isn't any solution for the "holes" problem but employment of other encoding techniques and efficiently organizing the file can reduce the impact of the problem on the storage requirements.

5. Additional work is required to make a realistic estimate of the storage space required for a GICS file implemented under S/360 FFS. To this end a plan was devised and is appended.

6. To date, approximately 1.5 man-weeks of effort have been expended and it is estimated it will require one man-week to implement the appended plan.

7. The effort outlined to estimate the S/360 FFS file size will require support from SG. Information and statistics must be collected, analyzed, and interpreted before a realistic storage estimate can be made. Specifics presently identified are:

A. The maximum field length for each field and the average length of variable (no size limit) fields.

B. The number of occurrences of each field in a file.

C. The valid set of characters for each field, e.g., numerics Ø thru 9 only, or twenty-six alphas, etc.

8. If the storage estimates are tolerable, I advise proceeding with the development of a FFT and implementation of one file using S/360 FFS for test purposes.

John J. Naughton

JOHN J. NAUGHTON
IBM

JJN:rmw

23 January 1970
ATTACHMENT

PLAN TO ESTIMATE S/360 FFS STORAGE REQUIREMENTS FOR GICS FILES

I Develop a preliminary file layout

A. Design the file structure - Determine the periodic, fixed, and variable fields. Specify logical grouping to develop periodic sets.

B. Determine structure adequacy - Insure that application retrievals and maintenance transactions are possible using preliminary design. Adjust structure accordingly.

C. Determine field types - For each field determine storage made--numeric, alpha, or alphameric.

D. Determine field encoding - Identify the fields to be encoded and the types of encoding to be used.

II Employ compaction techniques to reduce file size.

A. Identify fields for compaction - Determine those fields whose character set requires less than 2^8 binary combinations. Specify compaction for these fields.

III Collect statistics required for file size estimate

A. Determine average field length of variable fields - Run field analysis program if necessary to determine average field length.

B. Determine maximum field lengths - For each periodic and single valued field determine the maximum number of characters required.

C. Determine number of occurrences - For each field (fixed, periodic and variable) determine number of occurrences in file.

IV Refine field structure and compute total S/360 bytes required to store file.